

РАСПОЗНАВАНИЕ РУКОПИСНЫХ БУКВ С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ

Кулакова О.А., Воронова Л.И.

Московский Технический Университет Связи и Информатики

Москва, Россия

HANDWRITTEN LETTERS RECOGNITION USING NEURAL NETWORK

Kulakova O.A., Voronova L.I.

Moscow Technical University of Communications and Informatics

Moscow, Russia

В статье описывается модификация и применение программы для распознавания рукописных букв русского алфавита в рамках курсового проекта по дисциплине «Machine Learning. Обучающиеся технические системы», научный руководитель – д. ф.-м. н., проф. Л.И. Воронова.

Распознавание рукописного текста является важной задачей при обработке документов. Одним из самых распространенных методов машинного обучения при решении данной задачи являются нейронные сети.

Нейронные сети – это программная реализация сетей нервных клеток человека, т.е. биологических нейронных сетей. Активные работы над их созданием начались в 60-е годы, когда нейробиологами было установлено, что мозг – это миллиарды нейронов, соединенных между собой, а мозг человека способен выполнять сложнейшие задачи [1].

Основные понятия

Нейронная сеть состоит из связанных между собой нейронов, которые способны взаимодействовать друг с другом. Рассмотрим простую нейронную сеть, показанную на рисунке 1.

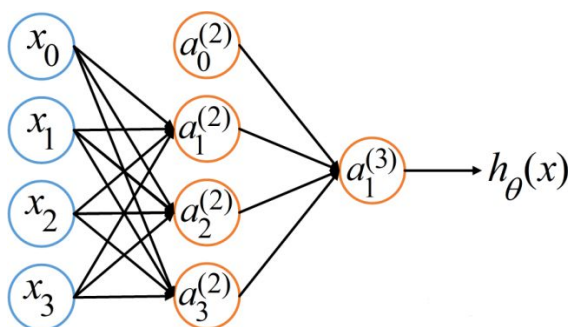


Рис.1. Модель нейронной сети

x_0, x_1, x_2, x_3 - это входные узлы, они относятся к первому слою, который называют входным; $a_0^{(2)}, a_1^{(2)}, a_2^{(2)}, a_3^{(2)}$ - обозначения для активации нейронов ($a_i^{(j)}$ - активация узла i в слое j), они относятся ко второму слою, который называют скрытым, количество скрытых слоев в сети выбирается при разработке архитектуры, именно в этом слое происходят основные вычисления; $a_1^{(3)} = h_\theta(x)$ - гипотеза. Последний слой называют выходным, т.к. он выдает конечное значение, вычисленное гипотезой, т.е. результат.

Логистическая регрессия – это самая распространенная функция активации, график которой представлен на рисунке 2:

$$h_\theta(x) = \frac{1}{1 + \exp(-\theta' X)},$$
 где θ – матрица параметров («весов»), X – матрица входных признаков [2].

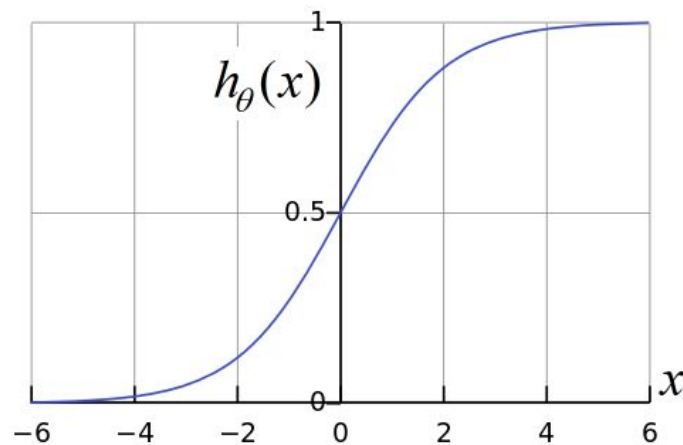


Рис. 2. Логистическая функция (сигмоид)

Важной задачей является обучение нейронной сети. При обучении происходит корректировка весов нейросети, таким образом сеть «учится» решать поставленные задачи [3]. Существует два вида обучения: с учителем и без. При обучении с учителем нейронной сети подается входной сигнал и желаемый результат, т.е. обучающая пара (x, y) , где x – входной сигнал, а y – выходной. На основании такого обучающего набора данных происходит изменение параметров сети. Наиболее часто для корректировки весов используется алгоритм обратного распространения ошибки. При обучении без учителя сеть обучается лишь на основе входных сигналов, желаемый результат остается неизвестным, она самостоятельно должна определить признаки, по отношению к которым параметры сети будут оптимизированы [4].

Распознавание образов

Сегодня технология компьютерного зрения активно применяется в медицине, робототехнике, сфере безопасности и во многих других областях. Задача распознавания образов относится к области компьютерного зрения. Нейронные сети активно применяют для решения задач распознавания образов.

Компания Google разработала нейронную сеть для распознавания рисунков. Для обучения сети они запустили онлайн-игру «Quick, Draw!», где пользователю ставится задача нарисовать какой-то предмет, в процессе прорисовки изображения компьютер пытается предугадать, что вы хотите нарисовать, таким образом создается обучающая пара: изображение и его название. При обучении учитывается не только сам рисунок, но и порядок прорисовки частей изображения [5]. На базе аналогичной нейронной сети существует Google переводчик рукописного текста, где ввод текста может производиться с помощью прорисовки текста на экране смартфона [6].

Распознавание рукописных букв

При создании программы для распознавания рукописного текста была модифицирована программа на базе нейронной сети для распознавания рукописных цифр, которая находится в открытом доступе и используется в качестве задания по программированию в курсе лекций по машинному обучению Andrew Ng от Стэнфордского университета [2].

Обучающий набор данных для нейросети целесообразнее взять готовый, т.к. в открытом доступе имеются базы знаний для обучения нейронных сетей. Мы использовали набор данных, состоящий из черно-белых изображений букв русского алфавита в виде файлов формата PNG, написанных различным почерком. Такой набор составлялся при помощи анкетирования, где респондентам надо было написать 29 букв русского алфавита своим обычным почерком. Изображения букв размерностью 145x145 пикселей преобразованы в матрицы при помощи функции `imread` и с помощью функции `reshape` из матрицы была сформирована матрица размером 1x21025, состоящая из нулей и единиц. Нули соответствуют белому фону, а единицы - черному. Использовалось 6090 изображений [7]. Далее все изображения были записаны в одну матрицу X размерностью 6090x21025. Примеры входных изображений представлены на рисунке 3. Обозначения классов, т.е. букв, представлены в таблице. На основании этих обозначений был сформирован вектор Y размерностью 6090x1, который состоит из чисел

от 1 до 29. Строка n вектора Y соответствует строке n матрицы X , т.е. каждой букве соответствует определенное изображение. X и Y были сохранены в файл расширением .mat.



Рис.3. Примеры входных изображений

Для распознавания изображений рукописных букв используется трехслойная нейронная сеть.

Входной слой нейронной сети имеет 21025 нейронов, так как в качестве обучающего набора используются изображения рукописных букв разрешением 145x145 пикселей, т.е. изображение состоит из 21025 пикселей. Таким образом входным сигналом для нейронной сети является каждый пиксель изображения.

Выходной слой имеет 29 нейронов, потому что используется 29 букв русского алфавита.

Для скрытого слоя важно выбрать оптимальное количество нейронов. Если выбрать слишком маленькое количество нейронов, то сеть может быть недообученной, при выборе слишком большого количества нейронов – переобученной. В таких случаях нейросеть не сможет правильно предсказывать значения. Можно воспользоваться правилом геометрической пирамиды [8], тогда следует выбрать 780 нейронов.

Соответствие букв и их обозначений

Класс (Буква)	Обозначение класса	Класс (Буква)	Обозначение класса
А	1	П	16
Б	2	Р	17
В	3	С	18
Г	4	Т	19
Д	5	У	20
Е	6	Ф	21
Ж	7	Х	22
З	8	Ц	23
И	9	Ч	24
Й	10	Ш	25
К	11	Щ	26
Л	12	Э	27
М	13	Ю	28
Н	14	Я	29
О	15		

Отрывок кода для определения переменных нейронной сети представлен на рисунке 4.

```
input_layer_size = 21025; %входной слой
hidden_layer_size = 780; %скрытый слой
num_labels = 29; %выходной слой
```

Рис.4. Определение переменных нейронной сети

Параметры сети были инициализированы случайным образом и сохранены в файле расширением .mat. Было сформировано две матрицы параметров: Theta1 и Theta2. Theta1 – это матрица весов для вычисления функций активации второго слоя сети, она имеет размерность 780x21026. Theta2 – это матрица весов для вычисления функций активации выходного слоя, т.е. гипотезы, она имеет размерность 29x781. Код для инициализации параметров представлен на рисунке 5.

```

init_epsilon = 0.016;
Theta1=rand(780, 21026)*(2*init_epsilon)-init_epsilon;
Theta2=rand(29, 781)*(2*init_epsilon-init_epsilon;

```

Рис. 5. Инициализация переменных

Для определения init_epsilon следует воспользоваться следующей формулой:

$$\varepsilon_{init} = \frac{\sqrt{6}}{\sqrt{L_{in} + L_{out}}}, \text{ где } L_{in} - \text{число нейронов во входном слое, } L_{out} -$$

число нейронов в выходном слое.

Переменные нейронной сети объявляются в основном файле, где так же производится вызов функций для реализации алгоритма прямого и обратного распространения ошибки, проверки градиента и обучения нейронной сети. Файлы расширением .mat загружаются с помощью команды load.

При реализации алгоритма прямого распространения последовательно рассчитываются функции активации. Код данного алгоритма представлен на рисунке 6.

```

a1 = [ones(m, 1) X];
z2 = a1*Theta1';
a2 = [ones(size(z2, 1), 1) sigmoid(z2)];
z3 = a2*Theta2';
a3 = sigmoid(z3);
h = a3;

```

Рис. 6. Алгоритм обратного распространения ошибки

Для реализации алгоритма обратного распространения [9] вычисляется ошибка активации узла σ :

$$\sigma_j^{(4)} = a_j^{(4)} - y_j$$

$$\sigma_j^{(3)} = (\Theta^{(3)})^T \sigma^{(4)} \cdot a_j^{(3)} \left(\begin{bmatrix} 1 \\ \dots \\ 1 \end{bmatrix} - a_j^{(3)} \right)$$

$$\sigma_j^{(2)} = (\Theta^{(2)})^T \sigma^{(3)} \cdot a_j^{(2)} \left(\begin{bmatrix} 1 \\ \dots \\ 1 \end{bmatrix} - a_j^{(2)} \right)$$

Для обучения нейронной сети используется усовершенствованный алгоритм по оптимизации [2], для минимизации дифференцируемой

функции. С помощью этой функции веса скорректируются, таким образом это уменьшит ошибку предсказания.

Эксперимент по распознаванию рукописных букв русского алфавита реализован с помощью программного обеспечения GNU Octave [10] – это свободно распространяемое программное обеспечение для математических вычислений, поддерживающее язык MATLAB [11].

Для обучения нейронной сети потребовалось около 8 часов, достигнута точность предсказания в распознавании букв в 85.6%.

При запуске программы для распознавания букв на экран программа будет выводить сообщение: «Neural Network Prediction: 26» и подаваемое на вход изображение, представленное на рисунке 7.

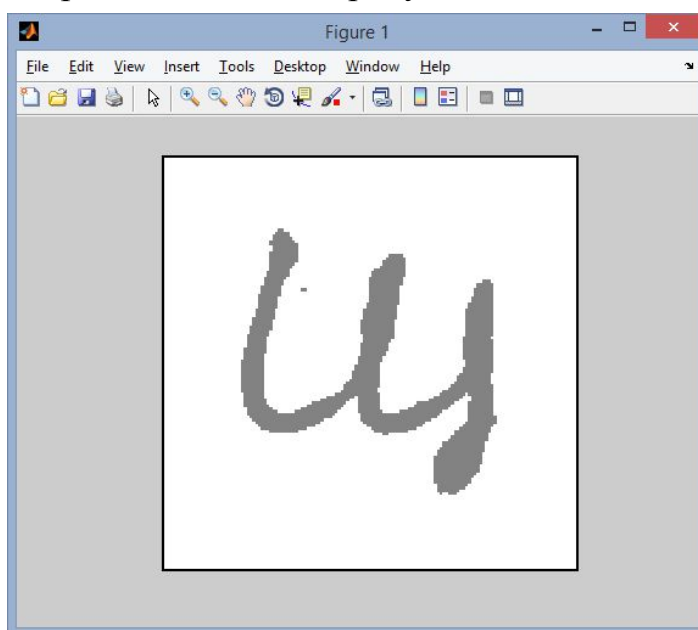


Рис. 7. Входное изображение.

Выводы

Распознавание образов - одно из активно развивающихся направлений интеллектуального анализа данных, при этом широко применяются нейронные сети. В статье рассмотрены основные понятия теории нейронных сетей, описана модификация программы для распознавания рукописных букв русского алфавита с помощью нейронной сети. В результате обучения нейронная сеть достигла точности предсказания в 85.6%.

Список источников и литературы

1. Е.Н. Горбачевская, С.С. Краснов. История развития нейронных сетей // Вестник Волжского университета имени В.Н. Татищева. 2015. № 1
2. Онлайн-курсы от ведущих университетов: <https://www.coursera.org>
3. В.М. Курейчик, Б.К. Лебедев, В.И. Божич. Эволюционная адаптация в обучении нейронных сетей // Известия Южного федерального университета. Технические науки. 1999. №3/том 13.
4. И.В. Воронов, Е.А. Политов, В.М. Ефременко. Обзор типов искусственных нейронных сетей и методов обучения // Вестник Кузбасского государственного технического университета. 2007. №3
5. Демонстрация алгоритмов машинного обучения от Google: <https://aiexperiments.withgoogle.com>
6. Google-переводчик: <https://translate.google.ru>
7. Эльмира Мустакимова. «Распознавание русского рукописного текста». Выпускная квалификационная работа. Национальный исследовательский университет «Высшая школа экономики», факультет гуманитарных наук. 2016. Москва
8. Медведев В.С. Нейронные сети MATLAB 6. — М.: ДИАЛОГ-МИФИ, 2002. — 496 с.
9. А.В. Горелова, Т.В. Любимова. Алгоритм обратного распространения // Наука и жизнь. 2015. № 38.
10. Официальный сайт Octave: <https://www.gnu.org/software/octave>
11. Официальный сайт MATLAB: <http://matlab.ru>